

UNITED STATES PATENT APPLICATION

OF

TERRY WEISSMAN

FOR

METHOD AND APPARATUS FOR SERIALIZING AN ASYNCHRONOUS
COMMUNICATION

PREPARED BY TELLME NETWORKS, INC.

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] This invention relates to the field of phone based communications. In particular, the invention relates to methods for providing serialized discussions from an asynchronous communication session.

Description of the Related Art

[0002] Conference calls (and chat lines) have previously allowed multiple participants into a single communication over the phone. On a conference call there are issues with people over-talking one another (e.g. if one person is talking the others must listen), there is minimal ability to review the communication after the fact, and the conversation must take place in real time. Thus if there is only one person in a particular call then that person will be quite bored—only able to talk to herself/himself. Similarly, if there are 100 people in a particular call almost no (meaningful) communication will be possible.

[0003] Accordingly, what is needed is a method and apparatus for handling an arbitrary number of users in a communication over a telephone interface by providing serialization of the asynchronous communication.

BRIEF DESCRIPTION OF THE FIGURES

[0004] Fig. 1 illustrates a logical view of participants in the serialized communication.

[0005] Fig. 2 illustrates the components of a phone application platform supporting the serialized communication.

[0006] Fig. 3 is a process flow diagram for participating in a serialized asynchronous communication.

[0007] Fig. 4 illustrates a logical view of participants in a conference-call style serialized asynchronous communication.

SUMMARY OF THE INVENTION

[0008] A method and apparatus for scalable handling of communications with varying numbers of participants over a telephone interface is described. The approach treats the different participants recorded communications as part of a larger asynchronous communication and provides a serializing (voice) user interface for participating in the conversation. This can be used to provide services ranging from phone-based discussion boards to more orderly teleconferences. Features may include moderation of comments, automatic removal of comments, and/or other features tailored to the specific use of the serializing approach.

DETAILED DESCRIPTION

A. Introduction

[0009] A method and apparatus for serializing an asynchronous communication is described. This approach can be used for a number of straightforward purposes from controlled phone conferencing to providing phone based chat lines that easily scale from 5 few to hundreds, or even thousands, of concurrent users. For example, in one embodiment of the invention, the approach is used to support a call for stock analysts. In another embodiment of the invention, the approach is used to support voice discussions about stocks and/or traffic information. End users of method and apparatus can use telephones, 10 including wireless telephones, to participate in the asynchronous communication.

[0010] The invention will be described in greater detail as follows. First, a number of definitions useful to understanding the invention are presented. Then, a presentation of the logical model for serializing the asynchronous communication is presented. Next, the basic architecture for a phone application platform supporting the method is presented. 15 Finally, the processes and features are presented in greater detail.

B. Definitions

1. Telephone Identifying Information

[0011] For the purposes of this application, the term telephone identifying information will be used to refer to ANI (automatic numbering identification) information, 20 CID (caller identification) information, and/or some other technique for automatically identifying the source of a call and/or other call setup information. For example, telephone identifying information may include a dialed number identification service (DNIS). Similarly, CID information may include text data including the subscriber's name and/or

address, e.g. "Jane Doe". Other examples of telephone identifying information might include the type of calling phone, e.g. wireless, pay phone, and/or hospital phone. Additionally, the telephone identifying information may include wireless carrier specific identifying information, e.g. location of wireless phone now, etc. Also, signaling system seven (SS7) information may be included in the telephone identifying information.

2. User Profile

[0012] A user profile is a collection of information about a particular user. The user profile typically includes collections of different information of relevance to the user, e.g., account number, name, contact information, user-id, default preferences, and the like.

10 Notably, the user profile contains a combination of explicitly made selections and implicitly made selections.

[0013] Explicitly made selections in the user profile stem from requests by the user to the system. For example, the user might add business news to the main topic list.

Typically, explicit selections come in the form of a voice, or touch-tone command, to save a particular location, e.g. "Remember this", "Bookmark it", "shortcut this", pound (#) key touch-tone, etc., or through adjustments to the user profile made through the web interface using a computer.

15 [0014] Additionally, the user profile provides a useful mechanism for associating telephone identifying information with a single user, or entity. For example, Jane Doe may have a home phone, a work phone, a cell phone, and/or some other telephones. Suitable telephone identifying information for each of those phones can be associated in a single profile for Jane. This allows the system to provide uniformity of customization to a single user, irrespective of where they are calling from.

[0015] In contrast, implicit selections come about through the conduct and behavior of the user. For example, if the user repeatedly asks for the weather in Palo Alto, California, the system may automatically provide the Palo Alto weather report without further prompting. In other embodiments, the user may be prompted to confirm the system's implicit choice, e.g. the system might prompt the user "Would you like me to include Palo Alto in the standard weather report from now on?"

[0016] Additionally, the system may allow the user to customize the system to meet her/his needs better. For example, the user may be allowed to control the verbosity of prompts, the dialect used, and/or other settings for the system. These customizations can be made either explicitly or implicitly. For example if the user is providing commands before most prompts are finished, the system could recognize that a less verbose set of prompts is needed and implicitly set the user's prompting preference to briefer prompts.

3. Topics and Content

[0017] A topic is any collection of similar content. Topics may be arranged hierarchically as well. For example, a topic might be business news, while subtopics might include stock quotes, market report, and analyst reports. Within a topic different types of content are available. For example, in the stock quotes subtopic, the content might include stock quotes. The distinction between topics and the content within the topics is primarily one of degree in that each topic, or subtopic, will usually contain several pieces of content.

4. Demographic and Psychographic Profiles

[0018] Both demographic profiles and psychographic profiles contain information relating to a user. Demographic profiles typically include factual information, e.g. age, gender, marital status, income, etc. Psychographic profiles typically include information about behaviors, e.g. fun loving, analytical, compassionate, fast reader, slow reader, etc.

As used in this application, the term demographic profile will be used to refer to both demographic and psychographic profiles.

5. Cookie

[0019] The term cookie, as used herein, refers to a structured data element

5 formatted according to the general principles of IETF RFC 2109 and/or some other state management standard.

A brief review of RFC 2109 may be useful. The core structure of a cookie is a name-value pair. The name is a token for identifying the cookie, e.g. "Customer", and the value is the value of that corresponding token, e.g. "Jane Doe".

10 Implicitly, each cookie is associated with the sending domain. According to RFC 2109, the implicitly set domain is the originating domain to which the HTTP request was sent. For example, if an HTTP GET request is sent to the request host "www.example.com", then the cookie set in response to that request would be implicitly associated with "www.example.com"

15 [0020] Additionally, a number of optional fields can be set, for example: a different domain for which the cookie is valid (Domain); a time to live (Max-Age); a version string (Version); etc. The phrases in parenthesis correspond to the RFC 2109 standard field names for the options.

C. Logical Model for Serializing an Asynchronous Communication

20 [0021] A logical model for understanding the method of serializing an asynchronous communication will be presented with reference to Figure 1 which illustrates a logical view of participants in the serialized communication. It is easier to understand the logical model if a particular example is selected.

[0022] The example we will use is a feature of a voice portal that allows users to discuss their thoughts on particular equity issues, colloquially: stocks. This feature would typically be associated with a stock quote type feature of the type available through the TELLME(TM) consumer service at +1 (800) 555-TELL(TM). It will be assumed that the asynchronous communication concerns the stock of WidgetCo, a fictitious company. Here is a sample dialogue that might lead a user to discover these communications features:

System: Hi there, welcome to Tellme...

User: Stock Quotes

System: Welcome to Tellme Stock Quotes, brought to you by <sponsor>...

User: WidgetCo

System: WidgetCo down five-eighths to forty-seven and a half...

System: Say 'Message Boards' to discuss WidgetCo stock with other users.

User: Message Boards

This brings us to the situation of Figure 1. In the example of Figure 1, there are four telephones (100A-D) coupled in communication with respective program execution threads (102A-D). Each thread of execution has a respective pointer (108A-D) to a place in an audio repository 104, the audio repository 104 including a plurality of (recorded audio) segments (110A-H).

[0023] Since our user just joined the communication we will assume she is the "last" phone in the picture, 100D. Notice that her program execution thread 102D has a pointer 108D to the earliest available segment 110A. (See below for a discussion of segment aging, starting positions and other details.) In contrast some of the other users (as represented by their phones) are listening to later points in the conversation as seen by their respective pointers to (chronologically) later recorded segments.

[0024] These recorded audio segments 110A-H correspond to the communication, each is a recorded message from one or more users (possibly even users that are no longer currently coupled in communication to the system by telephone). For example, the

segment 110A was recorded by a user no longer connected to the system. Here is a hypothetical transcription of the audio segments:

Segment 110A (user no longer still on phone): WidgetCo will easily make street estimates of sixty-cents a share.

5 Segment 110B (an earlier recorded segment whose user happens to still be connected via one of the phones 100A-D): I disagree with that, their Q4 balance sheet was hiding some nasty surprises.

Segment 110C (another user no longer on phone): Ditto, I think thirty cents a share will be lucky.

...

10 As this example shows, users can also record messages from within their thread of execution. Those messages are then appended to the end of the queue. For example, if as our example user of phone 100D listened to segment 110A, she vehemently disagreed she could immediately respond, e.g.:

System: (playing back segment 110A to phone 100D in program execution thread 102D)...

15 User: (interrupting) Respond

System: Please record your response, when you are done press the pound key.

User: (records message and hits #)

20 System: (optionally confirms or allows re-recording) (optionally estimates amount of time till a new caller will hear the message in the queue) (Allows user to control resumption of playback either where she left off, mid-segment, or at other points, e.g. resume, restart, next, previous, etc.)

The recorded message from phone 100D would, in this embodiment, be placed at the end of the queue, e.g. as segment 110I. That would mean that if the user of phone 100A finished the segment 110H she would be the first to hear the new comment in segment 110I, because in one embodiment users are automatically stepped through the queue in chronological order based on the time of the recording (oldest to newest).

Having presented the basic logical model for serializing an asynchronous communication, the basic architecture for supporting such a method and apparatus will now be described in greater detail.

30 D. Architecture

[0025] First, the hardware and software architecture of a system including an embodiment of the invention will be described with reference to Figures 2. Figure 2

illustrates a system including embodiments of the invention used to support phone applications, including the asynchronous communication application. The system of Figure 2 can be used to allow deployment of phone applications without the need for specialized hardware and/or software.

5 [0026] The following lists the elements of Figure 2 and describes their interconnections. Figure 2 includes the telephone 100A, a telephone network 204, a telephone gateway 207, a phone application platform 210, a VoiceXML browsers and supporting servers 212, a network 222, an application provider platform 220, a web servers 222, and the audio repository 104. The telephone 100A is coupled in
10 communication with the telephone network 104. The telephone network 204 is coupled in communication with the telephone gateway 207. The telephone gateway 207 is coupled in communication with the phone application platform 210. The network 222 is coupled in communication with the phone application platform 210 and the application provider platform 220.

15 [0027] The following describes each of the elements of Figure 2 in greater detail. The telephone 100A is a telephone interface to the phone application platform 210. The telephone 100A may be any sort of telephone and/or wireless telephone. For example the telephone 100A may be a land line phone, a PBX telephone, a satellite phone, a wireless telephone, and/or any other type of communication device capable of providing voice
20 communication and/or touch-tone signals over the telephone network 204. However, any audio signal carrying interface could be used.

[0028] The telephone network 204 may be the public switched telephone network (PSTN) and/or some other type of telephone network. For example, some embodiments of the invention may allow users with a voice over Internet Protocol (IP) phone to access the
25 phone application platform 110. The telephone network 204 is coupled to the telephone

gateway 207 that allows the voice communications and/or touch-tone signals from the telephone network 204 to reach the phone application platform 210 in usable form.

Similarly, the telephone gateway 207 allows audio signals generated by the phone application platform 210 to be sent over the telephone network 204 to respective

5 telephones, e.g. the telephone 100A. The telephone network 204 generally represents an audio signal carrying network.

July 92

[0029] The phone application platform 210 is comprised of one or more computers providing the VoiceXML browsers and supporting servers 212. (In this embodiment, VoiceXML is one of the implementation languages.) The particular configuration shown is designed support outsourced, or hosted, telephony provisioning as seen by the separation of the application provider platform 220 from the phone application platform 210. This allows the phone services to be provided by a different legal entity than the application. The implementation can be stored for access to the program via the web servers 222 using the HTTP protocol over the network 222 (the network 222 can be the Internet, a private network, an extranet, a virtual private network, or more generally any data carrying network). Similarly, the audio segments in the audio repository 104 can be accessed across the network using one or more protocols, e.g. HTTP, a streaming media protocol (e.g. RealAudio(TM)), etc. A more detailed description of one possible embodiment of the phone application platform 210 and features for working with audio content see United States Patent Application Serial No. 09/431,002, entitled "Streaming Content Over a Telephone Interface", having inventors Hadi Partovi, et. al., filed 01 Nov 1999, and assigned to the assignee of the current application.

[0030] Having described the basic architecture and some details, we now turn to implementation and other features in greater detail.

E. Implementation

[0031] The implementation will be described with reference to the process flow diagram of Figure 3 which illustrates how a user can participate in a serialized

asynchronous communication. Returning briefly to the logical view of Figure 1, the

5 program execution thread 102A-D loosely correspond to the state associated with the running applications (both VoiceXML and otherwise) on the phone application platform 210 and the application provider platform 220.

[0032] The process starts at step 310 with sign in. This step may be completely omitted, or delayed, in particular embodiments. For example, if this is offered as a free

10 feature then sign-ins may be unnecessary. In other embodiments, sign in is only required to access certain specified features, e.g. record a segment or set special options. In some embodiments, telephone identifying information is used to provide automatic sign-in without explicit user confirmation. As this process flow diagram suggests, there may be multiple conversations (each with a respective audio repository 104 or with the segments
15 within the audio repository 104 appropriately partitioned into conversations.)

[0033] Next, at step 320, the user can select a conversation to participate in. This may be automatically set, e.g. see example above where the conversation was selected implicitly/automatically because the user was in the stock board for WidgetCo. In some embodiments only a single conversation is supported. In one embodiment, each

20 conversation is assigned a (unique) number which can be manually entered by the user using DTMF or spoken voice commands. In another embodiment, each conversation is assigned a (unique) name which can be used by the user to access the conversation, e.g. by speaking the name. Additionally, if necessary there can be authentication for access to a particular conversation, e.g. based on user sign in, password, and/or some other

mechanism. In still other embodiments, the conversation is selected automatically based on telephone identifying information, e.g. dialed number identifies the conversation.

[0034] From a user perspective the meat of the process occurs next, at step 330, as segments from the conversation are played back, frequently in chronological order from oldest to most recent. The process can go a number of directions at step 330, the user can issue commands to:

1. control play back (stay on step 330, but within same or different segment)
2. record a response/segment (transition to step 332, and then back to 330)
3. change rooms (transition to step 320)
4. exit (transition out of Figure 3)
5. if there are no more messages (step 340) then either the user may exit out or be taken to hold sounds (step 342)

Each of these options will now be considered and explained in greater detail.

Play back controls are the most straightforward and may include commands for speeding or slowing the speed of playback (e.g. faster pace playback of segment at 1.5x recording speed), positioning within a segment (e.g., forward fast, rewind, etc.) and adjusting which segment is heard (e.g., skip ahead/back in audio repository). Typically, each of these commands may be associated with one or more DTMF inputs as well as one or more spoken word commands. In one embodiment, all segments are played back after use of a software program from Enounce Corporation, Palo Alto, California, <<http://www.enounce.com/>>, that shortens segments by reducing pauses between words and speeds the segment up without causing significant distortion.

[0035] Exiting the asynchronous communication is easily accomplished by providing an appropriate spoken or DTMF command, e.g. "exit", "menu", "goodbye", etc.

[0036] If a user reaches the end of the audio repository (no segments in ordinary playback order left to hear) then either hold sounds (step 342) are played or the process may terminate. The hold sounds may be music and/or any other sounds including help information and prompting for commands. If new messages arrive, in one embodiment, playback resumes automatically at step 330.

[0037] Whenever the user is listening to a segment (or after) she/he can reply to it. The basic approach is to transition to step 332 and accept audio input from the user (typically terminated with silence or a predetermined DTMF key press, e.g. pound) and then confirm the input before adding the segment.

[0038] This concludes the basic discussion of the features now it is helpful to talk about how this is scalable from one or two to hundreds of users as well as some advanced and alternative features.

[0039] Up until this point, the discussion has focused on what was assumed to be an ever-increasing audio repository for a conversation. However, scalability for handling both large and small volume conversations comes from purging (typically) older messages to maintain a certain relative overall audio repository/queue depth. Turning back to Figure 1, this would mean removing segment 110A (from the playback queue) after there were enough later entries. This queue shortening can be selectively applied on a per-conversation basis. Storage space permitting, it is not necessary to actually remove the recorded audio segment from the repository, just removing them from access and playback at step 330 is sufficient.

[0040] The queue shortening can work in several ways:

- remove segments left before a predetermined time (irrespective of queue size; useful for dated information, e.g. traffic conversation might get

cleared if older than 24 hours; or a “chat”-style conversation where all items only last a predetermined amount of time, e.g. one hour.)

- remove segments after there are enough subsequent audio segments to allow X additional minutes of participation in the conversation (this would mean that after, say 10 additional minutes worth of segments are recorded an earlier segment is removed; this helps control telecomm costs-especially useful in a free/low cost offerings)

These strategies are particularly well suited to handling both large volume and small volume conversations. For example, in a traffic related conversation, old information is of little use; however, the number of messages is somewhat less important. In contrast, for a heavy use conversation (e.g. about a popular stock or other topic), controlling queue length is important. The queue shortening techniques should balance controlling length against keeping a long enough queue (even of older segments) that callers can hear something even if there are not many current callers.

[0041] A slightly different approach is to switch the playback order, e.g. start at most recent and go to the oldest. Still other approaches are possible and may be adapted to the nature and purpose of a particular conversation. Also, multiple approaches can be applied to a single conversation in combination.

[0042] Other advanced features may include direct response to participants who recorded particular segments. In this embodiment sign ins (see step 310) would typically be used or (optionally) requested at message record time (within voice user interface of step 332). For example, after recording a segment the user might be prompted, “Would you like to receive direct responses to your segment, say yes or no.” For this example the relevant recorded segment is the segment 110A, recorded by a user of the phone 100A and the user requested direct responses.

[0043] Thus, when segment 110A is played back and another user hears the message the system may allow for a number of straightforward options: (1) record as normal and just forward the response to the message recorder; (2) provide different commands for public-vs-private responses, e.g. "Record" vs. "Private Message"; and/or combinations of these options. The specific manner in which the response is provided back to the message originator will now be discussed.

[0044] In one embodiment, certain users are allocated voicemail boxes to receive responses. In that case, the user can (without needing to reveal their personal phone number, or even their voicemail box number, to a responder) receive the responses in their voicemail. In one embodiment, having a voicemail box is a paid service available only to users who pay to access the services.

[0045] In other embodiments, the user can request outbound call notifications with the content of the reply. This could use the approach described in United States Patent Application No. 09/XXX,XXX, entitled "Interactive Callback Notifications", having inventors Michael J. Caferella, et. al., filed XX Jan 2001. In this configuration the user leaving a voice segment would (possibly implicitly from their sign-in) provide a calling number where they could be reached and after responses are recorded they would automatically be notified. One or more user interfaces (voice, web) may be provided to allow users to control notification times and numbers tried.

[0046] Some embodiments support both voicemail and outbound notification and the choice is selectable by users as they leave their recorded messages.

[0047] The queue segment feature also can be used as a type of moderation device. The simplest configuration of this is that a segment in the audio repository 104 is not played back to callers (during the process of Figure 3) until it is marked as approved. In one embodiment, one or more users (by way of their sign in) are identified as moderators

and provided access to the (as yet) unapproved segments during the process of Figure 3. After hearing a segment, the moderates can approve or disapprove the segment.

[0048] Similarly, in some embodiments, participants may be permitted to rate comments, e.g. 1-10 and off-topic. This could be used for small surveys but also to judge the popularity of a proposal or position without the need for a separate “polling”-type application.

[0049] A more complicated example showing moderation and the serialized asynchronous communication model will now be presented with reference to Figure 4. Additionally, this discussion will focus on the possibility of providing access to one or more “live” recorded audio segments in the sense that at least one segment is not fully recorded at the time access by others is allowed. Figure 4 illustrates a logical view of participants in a conference-call style serialized asynchronous communication. To flesh out the example, the phone 400E is being used by the discussion leader. In an analyst-style conference call this might be the CFO, CEO, etc. In this example, the analysts are called in on phones 100B-D and, an assistant to the company (possibly physically located near the users of the phone 400E) is reviewing analyst questions as the live portion is ongoing.

[0050] The live segment 410A is the audio (possibly not completely recorded at a particular time) from the conference call. For example, if the call starts at nine and ends up lasting till 10, then at 9:05 the live segment 410A will only be partially recorded.

However, the next day, the live segment 410A would be complete, etc.

[0051] As Figure 4 suggests, the analysts on phones 100B-D are listening to the presentation. The time-delay may depend on system load, but may also be live streaming. Because the analysts are listening to the streamed audio segment they can however pause, go back, or jump to the current front of the stream if they paused and want to pick up at what is being said now. Similarly, as analysts have questions they can say “Record”, or

similar, and leave their question. They can then either return to where they were in the stream or to the current front.

[0052] Meanwhile, the assistant using the phone 100A can review questions as they come in. This can allow them to be moderated, e.g. to remove objectionable and
5 duplicative material, to reorder the repository to place questions in rank order of important, and/or to provide the party placing the call (out of band) hints, e.g. by writing notes on a piece of paper, sending an Instant Message or email, or whispering in the other person's ear. For example if as the CFO is delivering the financials there are seven questions about an obscure line item, then the assistant could whisper "Hey, say something
10 about that tax credit", etc.

[0053] Handling of questions could then occur in a number of different possible fashions depending on the implementation preferences of the company and the style of dialogue preferred. In one embodiment, the remainder of the conference call occurs as described above with the CFO wading through the questions (segments 110B-110H) and
15 providing responses with those responses added as later segments. In one embodiment, the audio of a segment can be included as part of a response segment so that for example the audio of segment 110B could be included in the segment 110I (not shown) when the CFO responds to 110B. Note that if direct response to sender is enabled then the analyst could hang up and get her response without having to hold for hours.

20 [0054] In another embodiment, the person running the conference call (e.g. the CFO) presses a predetermined sequence (e.g. DTMF pound key) to terminate the initial live segment 410A. At that point the next approved segment, e.g. the segment 110B, is treated as a logical part of the stream for callers with them advancing to the segment 110B seamlessly. When the segment 110B finishes playing another live segment can be
25 automatically started for the CFO, e.g. the call leader. That new live segment would then

be heard by the analysts (on phones 100B-D) before they would hear the next question (in segment 110C). In simulated "live" variant, the analysts are not able to skip-ahead past the point in the conversation where the call leader is responding, e.g. to hear the questions or additional comments in segments 110C-H. However, after the whole call is over, e.g. call leader says we need to wrap this up, the entire collection of segments can be treated as described more generally above in connection with Figures 1 and 3.

[0055] This latter approach is particularly interesting in that it allows a high degree of moderation (in this example one assistant was used, but there could be multiple) and also allows analysts to be called back with the answers to their questions. Additionally, later on, analysts could leave follow ups to answers, etc.

[0056] There are a number of different models for financially supporting provision of these services. The most basic is an advertising supported model. In this configuration, users may be presented with ads prior to, during and after the process of Figure 3 to help compensate the service provider. Other mechanisms include per-minute charges (either to credit card, phone bill, ecash payment methods, etc.) and/or buffet style pricing at a flat amount per period of time. In one embodiment, even the paid service includes a limited amount of advertising. Other configurations may be totally user supported, e.g. conference call substitute type features of Figure 4 with costs based on number of minutes and/or number of participants. Additional features call outbound notifications and voicemail may be separately charged, e.g. \$1.00 to get a call back, etc.

F. Conclusion

[0057] In some embodiments, processes and apparatus of Figures 1-4 can be implemented using hardware based approaches, software based approaches, and/or a combination of the two. In some embodiments, the serialization of asynchronous

communications uses one or more computer programs that are included in one or more computer usable media such as CD-ROMs, floppy disks, or other media. In some embodiments, an execution thread, are included in one or more computer usable media.

[0058] Some embodiments of the invention are included in an electromagnetic

5 wave form. The electromagnetic waveform comprises information such as transcription generation programs, script handling programs, phonemic variation generation programs, script handling programs, and/or syllabication programs. The electromagnetic waveform may include the programs accessed over a network.

[0059] The foregoing description of various embodiments of the invention has

10 been presented for purposes of illustration and description. It is not intended to limit the invention to the precise forms disclosed. Many modifications and equivalent arrangements will be apparent.